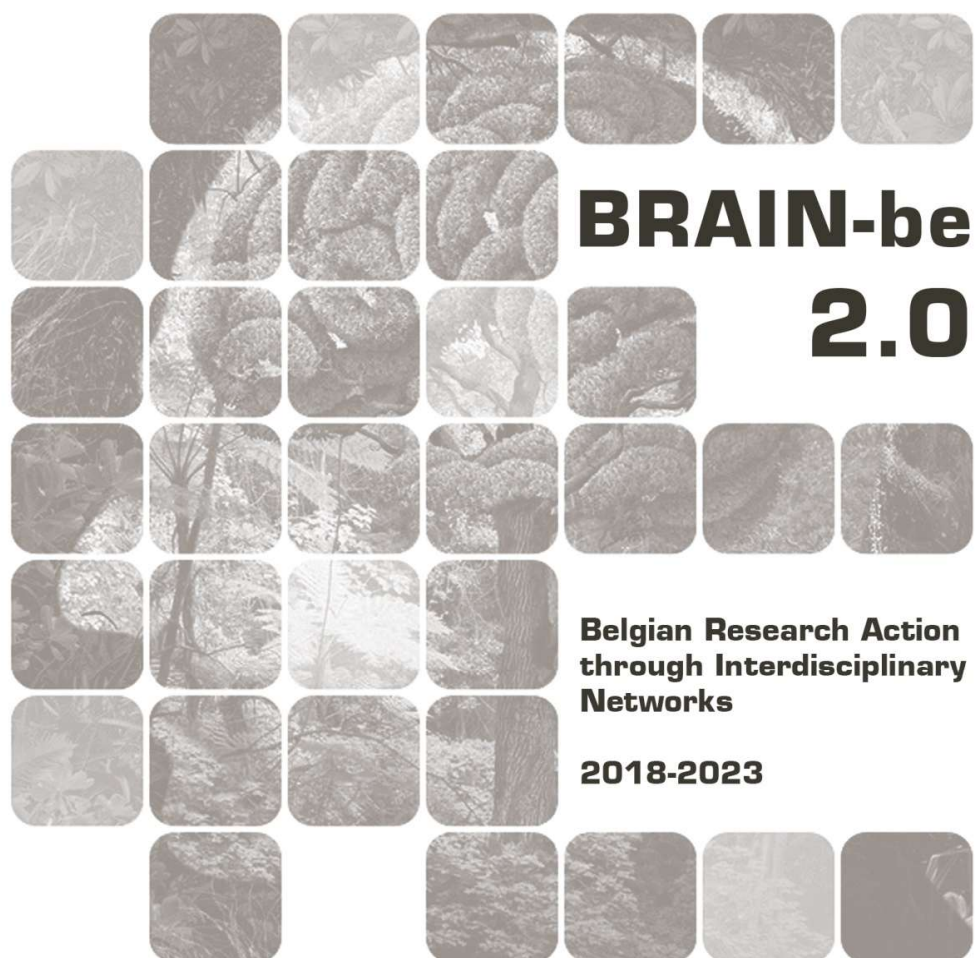


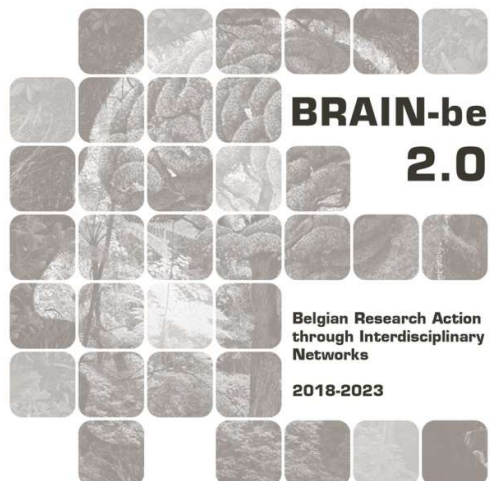


BelgicaWeb

Sustaining access to Belgium's born-digital heritage

Prof. dr. Julie Birkholz (UGhent – Ghent CDH), Prof. dr. Cécile DE TERWAGNE (UNamur - CRIDS), Friedel GEERAERT (KBR), George Caique GOUVEIA BARBOSA (UGhent – IDLab), dr. Isabelle GRIBOMONT (KBR), Élodie LECROART (UNamur - CRIDS), dr. Peter MECHANT (UGhent – mict), Fien MESSENS (UGhent – Ghent CDH), Alejandra MICHEL (UNamur – CRIDS), Kenzo MILLEVILLE (UGhent – IDLab), Christina VANDENDYCK (KBR), Prof. dr. Christophe VERBRUGGEN (UGhent – Ghent CDH), Prof. dr. Steven VERSTOCKT (UGhent – IDLab)

Pillar 2: Heritage science



NETWORK PROJECT

BelgicaWeb

Sustaining access to Belgium's born-digital heritage

Contract - B2/233/P2/BelgicaWeb

FINAL REPORT

PROMOTORS: Prof. dr. Cécile DE TERWAGNE (UNamur – CRIDS)
Prof. dr. Christophe VERBRUGGEN (UGhent – GhentCDH)
Friedel GEERAERT (KBR)

AUTHORS: Prof. dr. Julie Birkholz (UGhent – Ghent CDH)
George Caique GOUVEIA BARBOSA (UGhent – IDLab)
Dr. Isabelle GRIBOMONT (KBR)
Élodie LECROART (UNamur - CRIDS)
Dr. Peter MECHANT (UGhent – mict)
Fien MESSENS (UGhent – Ghent CDH)
Alejandra MICHEL (UNamur – CRIDS)
Kenzo MILLEVILLE (UGhent – IDLab)
Christina VANDENDYCK (KBR)
Prof. dr. Steven VERSTOCKT (UGhent – IDLab)





Published in 2026 by the Belgian Science Policy Office

Bolivar 30

Simon Bolivarlaan 30 bus 7

Boulevard Simon Bolivar 30 bte 7

B-1000 Brussels

Belgium

Tel: +32 (0)2 238 34 11

<http://www.belspo.be>

<http://www.belspo.be/brain-be>

Contact person: Corinne Lejour

Tel: +32 (0)2 238 34 91

Neither the Belgian Science Policy Office nor any person acting on behalf of the Belgian Science Policy Office is responsible for the use which might be made of the following information. The authors are responsible for the content.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without indicating the reference:

Birkholz, J., De Terwagne, C., Geeraert, F., Gouveia Barbosa, G. C., Gribomont, I., Lecroart, É., Mechant, P., Messens, F., Michel, A., Milleville, K., Vandendyck, C., Verbruggen, C., & Verstockt, S. BelgicaWeb: Sustaining Access to Belgium’s Born-Digital Heritage. Final Report. Brussels: Belgian Science Policy Office, 2026. 47 p. (BRAIN-be 2.0 - Belgian Research Action through Interdisciplinary Networks). Contract B2/233/P2/BelgicaWeb.

TABLE OF CONTENTS

ABSTRACT	5
CONTEXT	5
OBJECTIVES	5
CONCLUSIONS.....	5
KEYWORDS.....	5
1. INTRODUCTION	6
2. STATE OF THE ART AND OBJECTIVES	7
2.1 BORN-DIGITAL HERITAGE AND WEB ARCHIVES.....	7
2.2 ACCESS AND REUSE CHALLENGES	7
2.3 RESEARCH NEEDS AND USER-CENTRED DESIGN	7
2.4 PROJECT OBJECTIVES.....	8
3. METHODOLOGY	8
3.1 INTERDISCIPLINARY RESEARCH DESIGN	8
3.2 FOLLOW-UP COMMITTEE AND EXPERT PANEL	8
3.3 USER REQUIREMENTS RESEARCH	8
3.4 COLLECTION BUILDING AND PRESERVATION.....	9
3.5 TECHNICAL DEVELOPMENT	9
3.6 LEGAL RESEARCH.....	9
3.7 EVALUATION AND TESTING	9
4. SCIENTIFIC RESULTS AND RECOMMENDATIONS	9
4.1 INTRODUCTION	9
4.2 UNDERSTANDING USER REQUIREMENTS THROUGH CO-CREATION	10
4.3 CREATING THE BELGICAWEB COLLECTIONS: HARVESTING INFRASTRUCTURE, SEED LISTS AND QUALITY ASSURANCE.....	13
4.4 CREATING SUSTAINABLE ACCESS: ACCESS PLATFORM DEVELOPMENT, SEARCH INFRASTRUCTURE AND API DESIGN	18
4.5 DATA ENRICHMENT: TRANSFORMING WEB ARCHIVES INTO RESEARCH DATA	24
4.6 BUILDING A SUSTAINABLE LEGAL AND ETHICAL FRAMEWORK FOR ACCESS TO BORN-DIGITAL HERITAGE	29
4.7 SUSTAINABILITY, GOVERNANCE AND FUTURE PERSPECTIVES	35
4.8 SYNTHESIS: TOWARDS SUSTAINABLE ACCESS TO BELGIUM'S BORN-DIGITAL HERITAGE	37
5. DISSEMINATION AND VALORISATION	40
5.1 CONFERENCES ATTENDED.....	40
5.2 PRESENTATIONS AND WORKSHOPS ORGANIZED	40
5.3 PROJECT WEBSITES.....	40
5.4 PARTICIPATION IN PROFESSIONAL NETWORKS.....	41
5.5 INTERNAL DISSEMINATION AND INSTITUTIONAL IMPACT	41
5.6 OPEN SCIENCE AND REPOSITORIES	42
5.7 PUBLIC COMMUNICATION AND OUTREACH.....	42
5.8 CONCLUDING SYMPOSIUM	43
6. PUBLICATIONS	44
6.1 PEER-REVIEWED JOURNAL ARTICLES	44
6.2 CONFERENCE PRESENTATIONS AND PROCEEDINGS	44
6.3 DATASETS.....	45
6.4 TECHNICAL AND RESEARCH REPORTS	45
6.5 PROFESSIONAL PUBLICATIONS AND BLOG POSTS.....	47
7. ACKNOWLEDGEMENTS	47

ABSTRACT

Context

The Web has become one of the primary environments in which social, cultural, political and scientific life unfolds. Websites, blogs, online news platforms and social media contain valuable traces of contemporary society, yet much of this content is highly ephemeral and can easily disappear. Over the past decades, memory institutions, archives and libraries have increasingly recognized the importance of preserving born-digital heritage. While substantial progress has been made in collecting and preserving web content, significant challenges remain regarding access, reuse, legal compliance and long-term sustainability.

The Royal Library of Belgium (KBR), Ghent University and the University of Namur collaborated within the BELSPO-funded BelgicaWeb project **from 2024 to 2026**. Building upon the results of the [PROMISE](#) and [BESOCIAL](#) projects, BelgicaWeb aimed to open up Belgium's born-digital heritage by investigating sustainable approaches for collection building, access, enrichment, legal governance and long-term reuse of web and social media archives.

Objectives

The project pursued six main objectives:

- Investigating sustainable approaches for providing access to born-digital heritage collections.
- Creating and expanding collections of Belgian web and social media content.
- Developing methods for metadata enrichment and improved discoverability.
- Designing an API and access platform for researchers.
- Analyzing the legal frameworks governing access and reuse.
- Promoting awareness and use of born-digital heritage within Belgium and internationally.

The project adopted an interdisciplinary and user-centered methodology. Researcher and public needs were investigated through surveys, workshops, expert panels and participatory design approaches. Collection building activities were combined with technical experimentation and platform development. User requirements informed the implementation of an access platform and API (Application Programming Interface). Legal analyses addressed copyright, data protection, open data and artificial intelligence legislation. Throughout the project, iterative testing and evaluation ensured alignment between technical development and end-user expectations.

Conclusions

BelgicaWeb demonstrates that access to born-digital heritage requires more than preservation alone. Sustainable access depends on the integration of collection development, user-centred design, technical infrastructure, legal governance, metadata enhancement and long-term institutional commitment. The project resulted in a functional access platform and API, a prioritised set of researcher requirements, recommendations regarding responsible access and reuse, and a roadmap for future development of Belgian born-digital heritage infrastructures.

Keywords

Web Archiving, Born-Digital Heritage, FAIR Data, Access Infrastructure, Digital Humanities

1. INTRODUCTION

The web has become one of the primary spaces in which contemporary society documents itself. Public communication, political debate, cultural production, commercial activity and social interaction increasingly take place online. As a result, websites and social media platforms have become an important source of information for researchers across a wide variety of disciplines.

At the same time, online content remains highly vulnerable. Websites disappear, social media posts are deleted, platforms cease to exist and digital resources become inaccessible through technical or organizational changes. Web archiving has emerged as a means of preserving this information for future generations.

Over the past decade KBR has gradually developed expertise in preserving born-digital heritage through participation in initiatives such as [PROMISE](#) and [BESOCIAL](#). These projects focused respectively on web archiving and social media archiving and helped establish the foundations necessary for preserving Belgian online heritage.

The central ambition of the BelgicaWeb project was to make Belgium's born-digital heritage more Findable, Accessible, Interoperable and Reusable (FAIR). To achieve this objective, the project investigated the needs of researchers and other user groups, developed a prototype access platform and API, explored methods for metadata enrichment, and analysed the legal and ethical implications of opening up access to archived web content.

The project brought together expertise from multiple disciplines and institutions. KBR coordinated the project and contributed expertise in web archiving and digital preservation. Ghent University contributed expertise in digital humanities, user research and data enrichment. The University of Namur analysed the legal and regulatory frameworks governing access and reuse. Together, these partners developed an integrated approach that combines technological, legal and user-oriented perspectives.

This final report presents the main findings, achievements and recommendations of the BelgicaWeb project.

2. STATE OF THE ART AND OBJECTIVES

2.1 Born-Digital Heritage and Web Archives

The growing importance of digital culture has transformed the role of cultural heritage institutions. Websites, blogs, online communities and social media platforms increasingly constitute primary sources for understanding contemporary society. Consequently, web archiving has become an important activity for national libraries and archives worldwide.

Over the past twenty-five years, numerous national web archives have been established, including initiatives in Australia, the United Kingdom, France, Denmark, Luxembourg, New Zealand and many other countries. These initiatives have demonstrated both the opportunities and limitations of large-scale web preservation.

While preservation activities have matured considerably, access remains one of the main challenges. Many web archives remain difficult to discover, difficult to search and difficult to reuse as research data. Technical limitations, copyright restrictions and varying levels of digital literacy continue to affect user access.

2.2 Access and Reuse Challenges

The rapid growth of born-digital collections has created new opportunities for humanities and social science research. However, researchers often encounter barriers when attempting to work with archived web content.

Some require intuitive interfaces that support exploratory searching and browsing. Others require access at the data level through APIs and downloadable datasets. Researchers also increasingly expect enriched content, advanced search capabilities, visualization tools and interoperability with external datasets.

These developments require a shift from preservation-centred approaches towards user-centred access infrastructures.

2.3 Research Needs and User-Centred Design

Previous projects demonstrated that the needs of web archive users are highly diverse. Different user groups possess different levels of technical expertise, domain-specific knowledge and familiarity with web archives.

Understanding these needs therefore became a core objective of BelgicaWeb. Rather than designing technical solutions independently from users, the project adopted a participatory approach that integrated researchers, archivists, librarians and members of the public throughout the development process.

2.4 Project Objectives

The objectives of BelgicaWeb were organized around five interconnected themes:

- Researcher Needs and Collection Building
 - Develop collections and investigate user requirements regarding collection development and access.
- Access Platform and API
 - Develop a platform and application programming interface that enable access to archived web content.
- Data Enrichment
 - Explore enrichment methods including metadata enhancement, entity extraction and computational analysis.
- Legal and Ethical Frameworks
 - Analyse copyright, privacy, data protection and artificial intelligence regulation.
- Sustainability and FAIR Reuse
 - Develop recommendations supporting long-term preservation, reuse and institutional sustainability.

3. METHODOLOGY

3.1 Interdisciplinary Research Design

BelgicaWeb adopted an interdisciplinary methodology combining information science, computer science, digital humanities, archival science and legal research.

The project integrated technical development with stakeholder consultation to ensure that infrastructures reflected actual user requirements.

3.2 Follow-up Committee and Expert Panel

BelgicaWeb established a Follow-up Committee and an Expert Panel to support the project. The Follow-up Committee provided strategic advice and reviewed project progress and results. The Expert Panel played a more active role in the co-creation process by contributing to user requirements research, collection development, seed list creation, access platform evaluation and discussions on enrichment and quality assurance.

3.3 User Requirements Research

User requirements were investigated through surveys, workshops and participatory design activities. Workshops held during DH Benelux, RESAW and the network of Practice for Social Media Archiving provided valuable insights into researcher needs regarding selection, access and enrichment.

The collected requirements were prioritized using the MoSCoW methodology, allowing the project team to identify functionalities considered essential by users.

3.4 Collection Building and Preservation

Collection building activities included the development of selection policies, identification of seed lists and experimental crawls. The members of the follow-up committee submitted seed lists related to their respective research interest. In addition seed lists were created by the project team centrally. Browsertrix was selected as the primary harvesting infrastructure. Test collections were created and crawl reports were inspected to assess harvesting quality, completeness and usability.

3.5 Technical Development

The access platform was developed using [Symfony](#) and [API Platform](#). [SolrWayback](#) was integrated to support search and replay functionalities.

Development followed an iterative model in which user feedback informed continuous refinement of the access platform.

3.6 Legal Research

Legal analyses addressed copyright, privacy, personal data protection, text and data mining, open data policies and the [European Artificial Intelligence Act](#).

The legal work package translated these findings into practical recommendations for memory institutions seeking to provide access to born-digital heritage.

3.7 Evaluation and Testing

The access platform underwent expert review, usability assessment and beta testing.

Testing activities focused on accessibility, search functionality, metadata export, replay capabilities and overall user experience.

4. SCIENTIFIC RESULTS AND RECOMMENDATIONS

4.1 Introduction

The BelgicaWeb project builds upon a series of previous Belgian initiatives, most notably [PROMISE](#) and [BESOCIAL](#). These projects demonstrated both the importance and the complexity of preserving web and social media content. They provided valuable insights into legal constraints, collection development methodologies and researcher expectations. BelgicaWeb sought to extend these earlier efforts by focusing explicitly on access. The central question underlying the project was not merely how born-digital content can be collected, but how researchers and the public can meaningfully discover, explore, analyze and reuse such materials in the future.

A distinctive characteristic of BelgicaWeb was its commitment to a user-centred and evidence-driven design process. The project did not begin by developing technologies and subsequently searching for use cases. Instead, the development process was informed by extensive consultation with

researchers, archivists, librarians, heritage professionals and web archiving experts. User requirements were repeatedly collected and evaluated through workshops, surveys, expert panel consultations and usability evaluations. As a result, the project generated a detailed understanding of how different categories of users envision future interaction with web and social media archives.

The present chapter provides a comprehensive overview of the work undertaken within BelgicaWeb. It traces the project's development from the identification of user requirements to the construction of collections, the development of access mechanisms, the implementation of enrichment workflows and the examination of legal and governance challenges. Rather than presenting a simple chronology of deliverables, the report synthesizes the project's main findings and discusses the lessons learned while attempting to establish sustainable access to Belgium's born-digital heritage.

4.2 Understanding User Requirements Through Co-Creation

4.2.1 A User-Centred Research Strategy

A fundamental assumption underlying BelgicaWeb was that sustainable access infrastructures cannot be designed successfully without a detailed understanding of user needs. Web archives have historically been developed primarily from a technical preservation perspective. While considerable effort has been devoted internationally to capture technologies and storage infrastructures, considerably less attention has often been paid to the requirements of those who ultimately seek to use archived content.

To avoid this limitation, BelgicaWeb adopted a co-creation strategy from the outset. The objective was to involve users directly in the design and evaluation process and to ensure that technical development remained aligned with actual research needs.

Two consultative bodies played an important role throughout the project: the Follow-up Committee and the Expert Panel.

The Follow-up Committee provided strategic guidance throughout the project. Its members included experts in web archiving, digital preservation, research infrastructures, copyright and information policy. The committee acted as a sounding board for broader strategic questions and ensured that BelgicaWeb remained aligned with international developments in web archiving and digital heritage preservation. Discussions during Follow-up Committee meetings helped validate the overall project direction and provided opportunities to compare BelgicaWeb's approach with initiatives at other national libraries and heritage institutions.

Complementing this role, the project established an Expert Panel that became one of the key mechanisms for continuous user engagement and requirements gathering. By the end of September 2024, thirty-four experts had agreed to participate. The panel included researchers, archivists, librarians, heritage professionals, digital humanities specialists and web archiving practitioners.

Unlike the Follow-up Committee, which met periodically to discuss strategic matters, the Expert Panel was actively involved throughout the project and provided recurring feedback on concrete project outputs. Members participated in consultations concerning content selection, collection development, interface design, API functionality, enrichment priorities and quality assurance. Regular

communication through newsletters and online consultations ensured that the panel remained informed of project progress and able to contribute feedback at multiple stages of development.

The Expert Panel played a particularly important role in collection development. To support this activity, the project team developed a structured seed-list template through which panel members were invited to propose websites, online resources and thematic areas relevant to their fields of expertise. These contributions provided valuable input for collection creation and helped identify resources, communities and topics that might otherwise have remained underrepresented within the collections.

Similarly, the Expert Panel contributed substantially to the identification and validation of user requirements by filling out surveys, participating in workshops and in discussions regarding future functionalities of the access platform and API. The panel therefore served not simply as a consultation mechanism but as an active component of the co-design process.

The project complemented these expert consultations through a series of workshops and surveys intended to reach a broader community of stakeholders.

4.2.2 Workshops and Exploratory Consultation

BelgicaWeb organized a series of workshops to explore how researchers, heritage professionals and web archiving practitioners envision future access to born-digital collections. Rather than focusing on specific technical solutions, these workshops aimed to identify broader expectations regarding collection development, access and reuse.

The first workshop, organized during DH Benelux 2024, invited participants to reflect on the lifecycle of born-digital collections, from selection and preservation to access and interpretation. Discussions highlighted the importance of discoverability, transparency and contextual information. Participants stressed that web archives should function as research resources rather than merely preservation repositories.

A second workshop, organized within the Dutch-Flemish Network of Practice for Social Media Archiving, adopted a persona-based approach. Participants explored collection building, access and enrichment from the perspective of different user profiles and levels of expertise. This exercise highlighted the diversity of user expectations and the need for access environments that support both exploratory and advanced forms of research.

Across both workshops, several common themes emerged. Participants emphasized the importance of transparent collection development, rich metadata, flexible search and browsing functionality, and opportunities for computational access. There was also strong support for combining traditional consultation of archived webpages with mechanisms that enable dataset creation and large-scale analysis.

The workshops provided an important qualitative foundation for subsequent survey activities and helped shape the user-centered design approach adopted throughout the project.

4.2.3 Survey Results and Prioritization of User Requirements

To complement the workshop findings, BelgicaWeb conducted an international survey between October 2024 and January 2025. The survey targeted researchers, archivists, librarians, digital humanities scholars and heritage professionals. While not intended as a representative population study, it provided valuable insights into the expectations of current and potential users of web and social media archives.

The survey confirmed several themes that had already emerged during workshops and Expert Panel consultations. Respondents consistently emphasized the importance of transparency regarding collection development, advanced search and filtering capabilities, metadata access and opportunities for data export and computational reuse. Many participants viewed APIs and downloadable datasets as essential components of future web archive infrastructures.

The responses also demonstrated that users increasingly expect web archives to support multiple forms of research. While traditional search and browsing remain important, there is growing demand for data-oriented access that allows archived resources to be integrated into external analytical workflows.

Survey results were subsequently prioritised using the MoSCoW methodology. This exercise identified advanced search functionality, filtering options, metadata access, export capabilities, collection documentation and API access as the highest-priority requirements. These findings directly informed the development of the BelgicaWeb access platform, API and enrichment workflows.

4.2.4 Content Selection and Collection Development Requirements

The consultations undertaken throughout BelgicaWeb demonstrated that collection development and access design should be regarded as closely interconnected activities. Participants generally supported a combination of broad national web collections and more focused thematic or event-based collections. Rather than viewing these approaches as alternatives, they considered them complementary mechanisms for documenting contemporary society.

Respondents also highlighted the importance of preserving diverse content types, including textual content, images, audiovisual materials and downloadable resources. Particular attention was given to social media content, which was widely recognized as an essential but technically and legally challenging area of born-digital preservation.

Another recurring theme concerned documentation and transparency. Participants consistently indicated that future users require information about how and why content was selected, harvested and preserved. Such contextual information was considered essential for assessing representativeness and interpreting research results.

These findings informed both the collection-building activities described in the next section and the broader access strategy developed within BelgicaWeb. They reinforced the importance of

participatory collection development, transparent documentation practices and flexible access models capable of supporting diverse research needs.

Consolidated Key Findings and Recommendations

The co-creation activities undertaken throughout BelgicaWeb provided a consistent picture of current and future expectations regarding born-digital heritage.

Researchers increasingly view web and social media archives as research infrastructures rather than preservation systems alone. Users expect transparent collection development processes, advanced search and filtering capabilities, rich metadata and opportunities for computational access through APIs and reusable datasets.

The consultations also demonstrated broad support for combining large-scale national collections with more focused thematic collections. Participants highlighted the importance of preserving diverse content types, documenting collection decisions and providing contextual information that supports interpretation and reuse.

A final recurring theme was flexibility. Web archives serve users with different levels of expertise and diverse research needs. Future access environments should therefore combine intuitive interfaces for exploratory research with advanced services for computational and data-intensive analysis.

Key recommendations include:

- Maintain a user-centred and participatory approach to infrastructure development.
- Prioritize transparency regarding collection development and preservation processes.
- Support both traditional consultation and computational research workflows.
- Continue investing in metadata enrichment and contextualization.
- Combine broad collection strategies with curated thematic collections.
- Develop flexible access services that accommodate different user communities.

4.3 Creating the BelgicaWeb Collections: Harvesting Infrastructure, Seed Lists and Quality Assurance

The user requirements research provided strategic guidance for collection development, but collection creation required the translation of these requirements into practical workflows and preservation activities. This work package focused on establishing methodologies for identifying, harvesting and evaluating born-digital content suitable for long-term preservation and future access.

Collection development involved three interconnected components: the creation of seed lists, the evaluation of harvesting technologies and the quality assessment of archived collections. Together, these activities provided the foundation on which the BelgicaWeb access and enrichment infrastructures were built.

4.3.1 From Selection Criteria to Seed Lists

One of the first practical tasks of the project concerned the translation of broad collection development principles into operational seed lists.

Although consultations showed broad support for inclusive collection strategies, it quickly became apparent that practical constraints required carefully considered selection decisions. Available storage capacity, harvesting time, technical limitations and legal uncertainties all necessitated prioritization.

The project therefore adopted an iterative approach to seed selection. Initial seed lists were created using a combination of institutional expertise, previous experiences from PROMISE and BESOCIAL, and recommendations provided by project partners. These preliminary lists subsequently underwent several rounds of refinement and validation.

The Expert Panel played a particularly important role in this process. To facilitate structured input, the project team developed a dedicated template through which panel members were invited to propose candidate resources for inclusion in future collections. Experts were asked to identify websites, online resources and thematic areas relevant to their domains of expertise while providing contextual information to support subsequent collection development. This exercise transformed the Expert Panel from a purely advisory body into an active contributor to collection creation.

The submitted seed lists generated a valuable body of recommendations and provided insights into a wide range of topics, institutions, communities and online environments that might otherwise have remained underrepresented. In many cases, panel members went beyond simply identifying websites and provided explanations regarding the significance of specific resources, their relevance for future research and the rationale for their inclusion. These contributions enriched the project's understanding of the Belgian web landscape and facilitated a more informed prioritization process.

The Expert Panel's contributions also functioned as a mechanism for validating the representativeness of proposed collections. Through several rounds of review and discussion, the project team was able to identify gaps in coverage, refine thematic priorities and improve the overall diversity of the collections. The process demonstrated the value of combining institutional expertise with external domain knowledge when developing born-digital heritage collections.

Rather than relying exclusively on domain-based approaches, the project sought to combine multiple selection methodologies. These included thematic collections, event-based collections, institutional collections and broader sets of Belgian web resources. This multifaceted approach reflected the understanding that different research questions require different types of collections.

Several participants stressed the need to move beyond highly visible institutional websites and to also consider less prominent online resources. Community websites, activist platforms, cultural initiatives, local organizations and specialized thematic websites were frequently cited as important components of contemporary digital heritage that may otherwise be overlooked.

The resulting seed lists emerged from a combination of institutional expertise, project partner contributions and structured input from the Expert Panel. Through several rounds of review and refinement, the project developed collections that sought to balance representativeness, feasibility and research relevance while remaining responsive to the interests of potential future users.

4.3.2 Evaluating Harvesting Solutions

Once collection priorities had been identified, the project needed to determine how these resources could be harvested effectively.

A significant amount of effort was therefore devoted to evaluating harvesting technologies and workflows.

The project ultimately selected [Browsertrix Crawler](#) as its primary web harvesting solution. Several considerations motivated this decision.

First, Browsertrix is an open-source solution. Throughout BelgicaWeb, preference was consistently given to open-source technologies whenever suitable alternatives existed. This decision was motivated by long-term sustainability concerns, as well as by the desire to build upon tools that are actively maintained by an international community.

Second, Browsertrix provides support for modern web technologies. Many contemporary websites rely heavily on client-side rendering, JavaScript frameworks and dynamic loading mechanisms that are not always handled effectively by traditional crawlers. Browsertrix's browser-based approach offers significant advantages in this context.

Third, the access platform provides considerable flexibility regarding crawl configuration. Different collection types often require different capture strategies, and Browsertrix allows harvesting settings to be adapted accordingly.

Finally, the software benefits from an active development community and regular updates. This was viewed as an important consideration within a technological environment characterized by rapid change.

The adoption of Browsertrix followed a substantial testing phase. The project team undertook a series of experimental crawls using representative seed sets in order to evaluate capture quality, completeness, and replayability. The resulting collections were carefully examined to identify strengths and weaknesses and to determine whether the harvested content would be suitable for future access and reuse.

Particular attention was paid to the extent to which captured websites remained understandable and navigable after replay. Capturing content alone is insufficient if users are subsequently unable to reconstruct the original context or interact meaningfully with archived resources. The evaluation therefore considered both technical and user-oriented dimensions of quality.

The testing phase demonstrated that Browsertrix provides strong capabilities for harvesting contemporary websites. At the same time, it highlighted many of the limitations that remain characteristic of web archiving more generally. Dynamic content, externally hosted resources and rapidly changing web technologies continue to present challenges even for modern harvesting solutions.

As a result, the project adopted a pragmatic perspective that recognized preservation as a process of approximation rather than perfection. The objective was not necessarily to achieve flawless captures

in every case but rather to establish workflows capable of generating collections of sufficient quality to support future research.

4.3.3 Social Media Archiving as a Continuing Challenge

If harvesting websites remains technically demanding, social media archiving presents an even more complex set of challenges. The project was confronted with an environment in which major platforms increasingly restrict access to data and modify their technical interfaces, APIs and terms of service. These developments have significant implications for the long-term preservation of born-digital heritage.

Participants repeatedly identified social media content as a priority area for future collection development. Researchers are increasingly interested in studying online communities, political communication, public discourse and platform cultures. Preserving social media content is therefore essential if future generations are to understand contemporary digital society.

However, the collection of social media content is subject to additional technical, legal and organisational constraints. Several platforms have introduced restrictions on API access or significantly altered existing access conditions. Such changes can disrupt harvesting workflows and complicate long-term planning. The project therefore had to acknowledge that social media archiving remains highly dependent on decisions taken by platform owners and that access to content can change rapidly.

Additional difficulties arose from restrictions imposed on certain platforms within public sector environments. Such limitations created practical barriers for testing and collection activities and highlighted the extent to which heritage institutions often depend on infrastructures and services controlled by third parties.

These developments reinforced one of the recurrent conclusions of BelgicaWeb: sustainable access to born-digital heritage cannot rely exclusively on stable access to commercial platforms. Long-term preservation strategies must remain flexible and capable of adapting to changing technical, legal, and organizational environments.

Rather than presenting social media archiving as a solved problem, the project demonstrated the need for continued experimentation, monitoring and collaboration with the broader international web archiving community.

4.3.4 Quality Assurance and Collection Evaluation

Creating collections is only the first step. For collections to support meaningful research, it is equally important to ensure that they are usable, intelligible and sufficiently complete. Quality assurance therefore became an integral component of the collection development process.

The project approached quality from multiple perspectives. At a technical level, harvested collections were examined through the analysis of crawl reports to determine whether pages were captured successfully.

At a user level, quality was assessed through stakeholder feedback and subsequent usability evaluations. These activities helped identify opportunities for improving collection navigation, contextualization and documentation.

Once initial versions of the access platform became available, quality assurance activities also expanded towards user experience evaluation. A heuristic usability evaluation of the BelgicaWeb interface was conducted to assess the extent to which the access platform supported core usability principles. The evaluation indicated that the access platform generally performed well while simultaneously identifying areas for improvement concerning efficiency, scalability and support for certain user tasks. These findings informed subsequent refinements of the user interface and search environment.

Following the recommendations et remarks made on the heuristic usability evaluation, some changes were made:

- Bug fixing and removing unused interface elements, particularly in the dashboard.
- UX rework, where possible, in order to make the interface more intuitive, like help tooltips for forms, simplifying the menus and explanatory texts, indicators during heavy operations.
- Improvements to general error handling in forms.

The Expert Panel again played an important role in quality assurance activities. Members provided observations regarding collection coverage, highlighted potential blind spots and suggested areas for future improvement. Their involvement ensured that collection evaluation incorporated perspectives from multiple disciplines and user communities.

An important outcome of these activities was the recognition that quality in web archiving cannot be reduced to technical completeness alone. A collection may be technically successful yet still fail to meet user needs if it lacks contextual information, sufficient documentation or adequate coverage of the phenomenon under study. Consequently, BelgicaWeb adopted a broader understanding of quality encompassing technical, curatorial and user-oriented dimensions.

4.3.5 Lessons Learned from Collection Development

By the end of the project, several broader observations had emerged from the collection development activities. First, sustainable web archiving requires continuous interaction between technological and curatorial expertise. Neither harvesting technologies nor collection policies can be developed effectively in isolation.

Second, stakeholder involvement substantially improves collection quality. The recurring contributions made by the Expert Panel demonstrated the value of maintaining regular dialogue with potential users throughout collection development processes.

Third, flexibility remains essential. The rapid evolution of websites, platforms and online communication practices makes it impossible to define static collection strategies that remain valid indefinitely.

Fourth, collection development proved to be an iterative process. Collection priorities, harvesting configurations and evaluation criteria evolved throughout the project as new insights emerged and additional feedback became available.

Finally, collection development must be viewed as a long-term institutional commitment. Building born-digital collections is not a task that can be completed once and subsequently forgotten. Instead, it requires continuous adaptation, monitoring and reassessment.

The infrastructure and methodologies developed during BelgicaWeb provide an important foundation for this future work.

Key Findings and Recommendations

Several important conclusions emerged from the collection development and harvesting activities:

- Web archive collection development should be viewed as an ongoing and iterative process rather than a one-time activity.
- Structured contributions from the Expert Panel significantly improved collection quality, representativeness and research relevance.
- Providing experts with a seed list template proved an effective mechanism for gathering domain-specific knowledge and identifying content of potential long-term value.
- Broad collection strategies should be complemented by thematic and event-based collections.
- Open-source harvesting tools offer important advantages in terms of sustainability, transparency and flexibility.
- Social media archiving remains one of the most challenging areas of born-digital preservation and requires continued experimentation.
- Collection quality should be assessed from technical, curatorial and user-oriented perspectives.
- Documentation of collection decisions remains essential for supporting transparency and future research.
- Long-term success depends on maintaining both technical infrastructure and active stakeholder engagement.

4.4 Creating Sustainable Access: Access Platform Development, Search Infrastructure and API Design

If collection development formed the foundation of BelgicaWeb, the creation of sustainable access mechanisms represented the project's central ambition. Collecting and preserving born-digital heritage is only meaningful if researchers and other users can subsequently discover, understand and reuse the archived materials. Access was therefore not treated as a secondary activity that would follow collection building. Instead, it became a guiding principle throughout the project and influenced decisions relating to collection creation, metadata generation, access platform design and technical architecture.

The user requirements research demonstrated that contemporary researchers expect considerably more from web archives than the ability to consult individual archived webpages. Users increasingly expect access environments that support discovery, corpus creation, computational analysis and long-

term reuse. At the same time, accessibility, transparency and ease of use remain important requirements. The challenge for BelgicaWeb was therefore to create an access infrastructure capable of accommodating diverse research practices while remaining understandable to non-specialist audiences.

These different expectations presented an important challenge. The access infrastructure would need to support highly diverse research practices while remaining accessible to a broad audience. Consequently, the project adopted the principle that access should be layered rather than uniform. Different types of users should be able to interact with the collections in different ways, depending on their expertise, research questions and technical skills.

4.4.1 From User Requirements to Access Platform Design

The work undertaken in earlier phases of the project provided an important foundation for the access platform development. The workshops, survey results and Expert Panel consultations all demonstrated that users expected considerably more than a standard search interface. While search remained essential, participants repeatedly emphasized that they also required mechanisms for filtering, exporting, contextualizing and analyzing archived materials.

An important lesson emerging from the consultation process was that access needs to support the complete research lifecycle rather than merely the point at which a researcher retrieves a document. Researchers often begin with exploratory activities. They may not know precisely what exists within a collection and therefore need tools that allow them to identify relevant materials, understand collection structures and discover connections between resources. Once they have identified potentially useful materials, they frequently wish to narrow results, create corpora, export data and perform additional analysis. In many cases these research processes extend over weeks, months or even years.

The project therefore increasingly viewed access as a continuum rather than a single interaction between a user and an archive. This perspective influenced many of the design decisions that followed. Instead of focusing exclusively on retrieval, the development team sought to create an environment that could support multiple stages of research activity.

At the same time, several participants stressed that access should remain understandable to non-specialists. One of the recurring observations during workshops was that web archives can appear intimidating to users who are unfamiliar with harvesting technologies, [WARC](#) files or snapshot-based preservation practices. Providing contextual information, collection descriptions and user-friendly navigation mechanisms therefore became an important objective.

These observations reinforced the importance of a user-centered design approach. Rather than simply exposing archived content through existing tools, the project sought to build access mechanisms around the needs identified by future users.

4.4.2 Developing the BelgicaWeb Access Platform

The development of the BelgicaWeb access platform represented one of the most visible outcomes of the project. The access platform was conceived as the primary point of interaction between users

and the archive and was therefore designed to accommodate different levels of expertise and different forms of engagement with born-digital heritage.

Rather than developing a simple search portal, the project envisioned an environment combining discovery, research and collection management functionalities. This broader vision reflected the understanding that web archives increasingly operate as infrastructures supporting a variety of scholarly and cultural heritage activities.

The access platform architecture evolved around three interconnected environments. The first environment focuses on public access. This layer provides entry points for visitors who wish to explore collections, browse archived materials and familiarize themselves with the archive. Particular attention was devoted to ensuring that users can understand what collections exist, how they were created and how they can be explored. Collection descriptions, thematic overviews and contextual information therefore became important components of the user experience.

A second environment focuses on registered users and longer-term research activities. Researchers frequently return to collections repeatedly during the course of a project. They often perform multiple searches, compare resources across different visits and gradually build corpora that form the basis of more detailed analysis. The access platform was therefore designed with personalized functionalities in mind, including the possibility of saving searches, bookmarking resources and maintaining a personalized working environment.

A third environment addresses the needs of collection managers and administrators. Behind every access platform lies a substantial amount of ongoing maintenance, metadata management and collection administration. The project therefore considered administrative functionality an essential component of sustainability rather than an afterthought. Effective access requires the ability to manage collections, update information and monitor the overall health of the infrastructure.

Taken together, these three environments reflect a broader understanding of access. Rather than treating access solely as a matter of retrieval, BelgicaWeb increasingly approached it as an ecosystem within which different categories of users interact with collections in different ways.

4.4.3 Building a Sustainable Technical Architecture

The technical architecture underpinning the access platform was guided by sustainability considerations from the outset. As with other components of the project, there was a strong preference for open-source solutions that could be maintained, adapted and extended beyond the duration of BelgicaWeb.

The decision to build the access platform using Symfony must be understood in this context. Symfony provides a mature and widely adopted development framework that combines flexibility with long-term stability. Its extensive developer community, modular architecture and strong documentation make it particularly suitable for projects that seek continuity after the end of dedicated project funding.

Equally important was the framework's ability to support interoperability. BelgicaWeb was never envisioned as a standalone application operating in isolation. Instead, it forms part of a broader ecosystem that includes harvesting tools, indexing infrastructures, APIs and enrichment services. A

flexible development framework was therefore essential for ensuring that these different components could be integrated effectively.

The use of open-source technologies also aligns with broader strategic objectives within the heritage sector. By relying on technologies that can be inspected, adapted and maintained by multiple institutions, the project reduces dependence on proprietary solutions and contributes to long-term sustainability.

4.4.4 Search and Discovery

Search functionality emerged as one of the most consistently discussed topics throughout the project. This is perhaps unsurprising: for many users, search constitutes the primary mechanism through which web archives become accessible.

However, the consultations conducted during BelgicaWeb demonstrated that expectations regarding search have expanded significantly. Researchers no longer simply wish to find specific webpages. They increasingly seek tools that enable the exploration of large collections, the identification of patterns and the discovery of relationships between resources.

This distinction between retrieval and discovery became an important theme during development. Retrieval involves locating known resources. A researcher may, for example, wish to find a specific website, URL or page captured at a particular moment in time. Discovery is different. Here, the objective is to explore a collection, identify trends or uncover resources that were not previously known.

Supporting both approaches requires a diverse set of search functionalities. Consequently, the project explored a combination of keyword searching, URL searching, collection-level searching, metadata searching and temporal filtering. Researchers repeatedly emphasised the importance of combining multiple search criteria, particularly when working with large collections spanning extended periods of time.

Filtering functionality received particular attention. Large result sets are often difficult to navigate and can quickly become overwhelming. The ability to narrow results progressively through dates, domains, collections or metadata fields was therefore regarded as a fundamental requirement. The project increasingly viewed filtering not as a supplementary feature but as a core component of the search experience.

Multilingualism represents an additional challenge within the Belgian context. Collections contain content created in multiple languages and future users will inevitably approach these materials using different linguistic perspectives. Although multilingual retrieval remains an evolving area, the project identified it as an important direction for future development and investigation.

4.4.5 Search Infrastructure and SolrWayback

Behind the user interface lies a complex infrastructure responsible for indexing and retrieving archived content. Effective access depends not only on interface design but equally on the quality and flexibility of the underlying search environment.

For this reason, the project explored the potential of SolrWayback as a core access technology. SolrWayback has become one of the most widely adopted solutions within the international web archiving community and has been implemented successfully by several national libraries and heritage institutions.

Its appeal lies in its capacity to bridge the gap between preservation and access. SolrWayback combines indexing, retrieval and replay functionality within a single environment, making it possible to move seamlessly between search results and archived content. At the same time, it provides considerable flexibility regarding search configuration and metadata integration.

An important aspect concerned the indexing of archived resources. Effective access depends on the ability to transform harvested content into searchable collections through the creation and maintenance of indexes. The project therefore investigated how metadata, full-text content and crawl information could be structured and indexed in order to support efficient retrieval, filtering and discovery across large collections. These activities highlighted the close relationship between collection quality, metadata quality and search performance.

The evaluation and implementation activities undertaken during BelgicaWeb confirmed that SolrWayback offers a strong foundation for future access services. At the same time, the project demonstrated that successful implementation requires careful consideration of local requirements, metadata structures and user expectations.

The resulting infrastructure should not be viewed as a fixed technical solution. Instead, it represents a foundation upon which additional services and functionalities can be developed in the future.

4.4.6 API Development and Computational Access

Perhaps one of the clearest indications of changing researcher expectations concerns the growing demand for programmatic access. Throughout the consultation activities, researchers repeatedly stressed their desire to interact with archival content as data rather than solely as documents.

This trend reflects broader developments within digital humanities, computational social sciences and data-intensive research. Increasingly, researchers seek to create datasets, perform automated analyses and develop reproducible workflows. Such activities cannot easily be accommodated through graphical interfaces alone.

The project therefore initiated the development of a dedicated API intended to complement the user-facing access platform. Rather than replacing the graphical interface, the API provides an additional layer of access designed specifically for computational use.

Several categories of functionality were identified during the design process. These include the retrieval of archived resources, metadata access, search functionalities and export mechanisms. Together, these services create opportunities for researchers to integrate archived content directly into external analytical environments.

The development of the API also reflects a broader conceptual shift. Historically, web archives primarily delivered information through interfaces designed for human consultation. Contemporary research infrastructures increasingly need to support communication between systems as well as

between institutions and users. APIs therefore play an important role in ensuring interoperability and facilitating new forms of scholarly reuse.

For BelgicaWeb, the API should be understood not merely as a technical component but as a strategic investment in future research possibilities.

4.4.7 Evaluation and Iterative Improvement

Throughout the project, access platform development was treated as an iterative process rather than a linear sequence of implementation steps.

Usability evaluations formed an important part of this approach. A heuristic assessment of the access platform identified both strengths and opportunities for improvement. The results were generally positive and indicated that the access platform provides a sound basis for future development. Nevertheless, the evaluation also highlighted areas where navigation structures, information hierarchies and support for specific research tasks could be further refined.

Importantly, these evaluations reinforced one of the central lessons emerging throughout BelgicaWeb: successful access infrastructures depend as much on user experience as on technical functionality. Powerful search systems and extensive collections provide limited value if users struggle to understand how they should be used.

The project therefore incorporated feedback and evaluation throughout the development process, ensuring that technical implementation remained closely connected to user requirements.

4.4.8 Towards a Research Infrastructure for Born-Digital Heritage

Taken together, the access platform, search infrastructure and API activities reveal a broader transformation in how web archives are conceptualised.

Traditionally, access has often been viewed as a service layer added after preservation activities were completed. BelgicaWeb suggests a different perspective. Access influences collection development, metadata creation, enrichment priorities and technical architecture from the earliest stages of the archival lifecycle.

Users increasingly expect web archives to function as research infrastructures capable of supporting discovery, analysis and reuse. The developments undertaken during BelgicaWeb therefore represent more than a set of technical outputs. They contribute to a broader vision in which born-digital heritage collections become active components of scholarly and cultural ecosystems.

Key Findings and Recommendations

- Access should be regarded as a central component of web archiving rather than a service added after preservation has taken place.
- Researchers increasingly expect web archives to operate as research infrastructures supporting multiple stages of the research process.
- Access platform design should accommodate both novice and expert users through layered forms of access.
- Search environments need to support both precise retrieval and exploratory discovery.

- Open-source technologies provide important advantages in terms of sustainability, transparency and future development.
- APIs are becoming essential tools for enabling computational research and supporting interoperability.
- User experience and usability should be considered strategic concerns rather than purely technical issues.
- Future developments should continue to integrate graphical interfaces, APIs and enrichment services within a unified access ecosystem.

4.5 Data Enrichment: Transforming Web Archives into Research Data

While the development of collections and access mechanisms constituted two essential pillars of BelgicaWeb, a third pillar emerged from the recognition that access alone is not sufficient. Contemporary web archives contain enormous volumes of heterogeneous material that can be difficult to navigate, interpret and analyse. Even where content has been successfully preserved and indexed, researchers are often confronted with challenges relating to scale, discoverability and contextualisation. The project therefore devoted substantial effort to the development of data enrichment workflows intended to transform archived web content into richer and more usable research resources.

The importance of enrichment became apparent throughout virtually every stage of the project. During workshops, consultations and survey activities, researchers repeatedly highlighted the difficulties of working with large collections of archived web content. Traditional metadata often provides only limited information regarding the substance of a webpage or the relationships between different resources. Researchers frequently expressed a desire for higher-level descriptions and analytical metadata that would allow them to explore collections more efficiently and formulate research questions more effectively.

These observations reflect a broader development within digital cultural heritage. As collections continue to grow, the challenge increasingly shifts from preservation itself towards understanding what has been preserved. Data enrichment can therefore be viewed as an important bridge between collection building and research use. By adding additional layers of information to archived resources, enrichment makes collections more discoverable, more interpretable and ultimately more reusable.

Within BelgicaWeb, enrichment was never conceived as a replacement for traditional archival description. Instead, it was understood as a complementary layer that could support researchers in navigating increasingly complex collections. The project sought to explore how a combination of metadata generation, quality assessment, geolocation, copyright analysis and artificial intelligence could make born-digital heritage more accessible for different user communities.

4.5.1 Understanding the Need for Enrichment

The need for enhanced metadata emerged particularly strongly from the user consultation process. Researchers consistently indicated that they required more than technical capture information and basic descriptive metadata. Many participants described situations in which potentially relevant

resources remained difficult to identify because the available descriptions provided little indication of the content contained within archived webpages.

This challenge becomes increasingly pronounced as collections grow larger. A researcher working with a handful of archived websites may still be able to manually inspect individual resources. However, when collections contain thousands or even millions of captured pages, manual exploration quickly becomes impractical. Researchers therefore require mechanisms that help them identify relevant subsets of material and understand collection contents at a higher level of abstraction.

The consultations further demonstrated that users increasingly expect web archives to support both exploratory and analytical forms of research. Exploratory users often wish to obtain rapid overviews of collection contents, identify important themes or discover previously unknown resources. Analytical users may seek structured information that can support quantitative analysis, corpus creation or computational processing. Both groups benefit from richer forms of metadata and contextual information.

Another recurring theme concerned trust and transparency. Researchers frequently expressed a desire to understand the quality and reliability of archived content. Questions regarding completeness, capture success and preservation quality arose repeatedly throughout discussions. This highlighted the importance of developing quality indicators capable of informing future users about the characteristics and limitations of archival resources.

Taken together, these observations suggested that enrichment should address not only the content of archived webpages but also the broader context within which these resources were collected and preserved.

4.5.2 Developing the BelgicaWeb Enrichment Pipeline

To address these needs, BelgicaWeb developed an enrichment infrastructure capable of processing archived web collections and generating additional metadata layers.

The enrichment workflow was designed around [WARC](#) files, which remain the standard preservation format for web archives. Rather than treating these files solely as preservation objects, the project explored how they could also serve as inputs for metadata generation and analytical processing.

The resulting workflow processes a single WARC file and produces a structured metadata record describing the archived content. By repeating this process across larger collections, it becomes possible to generate datasets capable of supporting both human interpretation and computational analysis.

An important design principle throughout development was modularity. Different enrichment components address different dimensions of archived content and can be combined depending on the needs of specific collections or research scenarios. This approach provides flexibility while simultaneously facilitating future extensions.

The enrichment workflow also reflects a broader shift in how web archives are conceptualized. Traditionally, archived content has often been viewed primarily as material to be preserved and

consulted. Within BelgicaWeb, archived content increasingly became viewed as data that can be processed, analyzed and enriched in order to generate additional forms of value.

This perspective aligns closely with the project's broader commitment to [FAIR principles](#). Richer metadata contributes directly to the findability, accessibility, interoperability and reusability of archived collections.

4.5.3 Quality Assessment as Metadata

One of the first enrichment components developed within the project concerns quality assessment.

Quality represents a recurring concern within web archiving. Researchers need to understand whether a capture was successful, whether important content elements are missing and how representative a particular archived resource may be. Yet such information is not always immediately visible within traditional web archive interfaces.

The BelgicaWeb enrichment workflow therefore incorporates the generation of quality-related metadata.

Rather than providing a simple binary distinction between successful and unsuccessful captures, the project explored a broader range of indicators capable of describing different dimensions of archival quality. These indicators include measures relating to HTTP response success, extraction rates, content availability, domain diversity and temporal coverage.

The value of these metrics extends beyond purely technical evaluation. For researchers, they provide contextual information that can help interpret analytical results and identify potential limitations within datasets. For collection managers, they offer mechanisms for monitoring collection health and identifying preservation issues requiring further attention.

Quality assessment therefore serves multiple audiences simultaneously. It supports preservation activities while also contributing to transparency and research reuse.

More broadly, the inclusion of quality information reflects one of the central lessons emerging from BelgicaWeb: access should not merely expose archived content but should also provide users with sufficient information to evaluate that content critically.

4.5.4 Geographical Context and Domain-Level Analysis

A second component of the enrichment workflow focuses on geographical information.

Web archives often contain resources originating from diverse technical and geographical contexts. Although archived webpages may appear as coherent collections, they frequently depend upon infrastructures distributed across multiple countries and hosting environments.

To support geographical analysis, the project developed functionality capable of generating metadata concerning hostnames, IP addresses and geographical locations associated with archived resources.

This enrichment layer creates several opportunities. First, it allows collection managers to gain a better understanding of the geographical distribution of preserved resources. Second, it introduces

new possibilities for researchers interested in spatial patterns within web collections. Finally, it contributes to the broader contextualization of archived resources by linking content to infrastructure.

Although geographical metadata should not be interpreted as a direct indicator of cultural or national identity, it nevertheless adds an additional analytical dimension that may prove valuable for future research.

The development of geolocation functionality also reflects the project's commitment to generating metadata that supports a wide range of potential use cases rather than serving a single predefined research agenda.

4.5.5 Copyright Indicators and Rights Awareness

One of the most distinctive aspects of the BelgicaWeb enrichment workflow concerns the integration of copyright-related analysis.

The legal studies undertaken throughout the project demonstrated that copyright remains one of the most important factors shaping access to born-digital heritage. Decisions regarding preservation, access and reuse are often influenced by the copyright status of archived materials.

The enrichment infrastructure therefore explores mechanisms for identifying indicators that may suggest the presence of copyright-related information within archived content.

This process does not replace legal analysis, nor does it provide definitive determinations regarding rights status. Instead, it functions as an additional layer of information that can support subsequent assessment and decision-making.

The inclusion of copyright indicators reflects a broader philosophy that legal considerations should be integrated into archival workflows wherever possible rather than treated as entirely separate processes. By embedding rights awareness within metadata generation activities, the project sought to create foundations for more informed access management in the future.

This approach also demonstrates how technical and legal work packages within BelgicaWeb increasingly intersected. Questions initially explored through legal analysis ultimately influenced technical design choices and metadata development priorities.

4.5.6 Artificial Intelligence and Automated Metadata Generation

Perhaps the most innovative component of the enrichment workflow concerns the integration of artificial intelligence techniques.

Throughout the project, considerable attention was devoted to exploring how large language models and related AI technologies might assist in the creation of richer descriptive metadata. These investigations were motivated partly by user requirements and partly by broader developments within the cultural heritage sector.

Researchers repeatedly indicated that they would benefit from metadata capable of describing content at a conceptual level. Traditional metadata often captures technical characteristics or

bibliographic information but provides limited insight into themes, entities, topics or narratives present within archived resources.

Artificial intelligence offers new possibilities in this regard.

Within BelgicaWeb, AI-supported enrichment focused on several categories of metadata generation. These include language identification, sentiment analysis, thematic categorization, topic extraction, named entity recognition and automated summary creation. Together, these functionalities seek to provide researchers with higher-level descriptions of archived content and facilitate more efficient exploration of collections.

Named entity recognition is particularly promising from a research perspective. By identifying references to persons, organizations, locations and other entities, it becomes possible to reveal patterns and relationships that may remain difficult to detect through manual inspection alone.

Similarly, automated summaries can provide users with rapid insights into the content of archived resources without requiring them to inspect large numbers of webpages individually. Such functionality may prove particularly valuable during exploratory stages of research.

The project also explored multimodal approaches that combine textual and visual analysis. This included the generation of image descriptions based on screenshots captured during the replay process. Such approaches acknowledge the increasingly visual nature of contemporary web communication and seek to create metadata capable of describing both textual and visual dimensions of archived content.

These experiments demonstrated that computer vision and multimodal AI models may help researchers navigate visually rich collections by generating descriptive metadata for screenshots and other visual resources. Such approaches may become increasingly valuable as visual communication continues to occupy a larger role within contemporary web culture.

4.5.7 Opportunities, Limitations and Responsible Use

Although the use of AI creates significant opportunities, the project consistently adopted a cautious and critical perspective regarding automated enrichment.

A recurring concern throughout discussions involved transparency. Researchers need to understand how metadata has been generated and what limitations may affect the resulting outputs. Automated descriptions should therefore be viewed as supplementary metadata rather than authoritative interpretations.

Questions regarding reproducibility, reliability and bias also emerged repeatedly. AI systems can generate inaccuracies, oversimplifications or inconsistent results. Consequently, the project did not approach AI as a mechanism for replacing existing archival practices. Instead, it explored how such technologies might assist archivists and researchers while leaving ultimate interpretation in human hands.

This perspective aligns closely with broader discussions regarding responsible AI within cultural heritage institutions. The value of automated enrichment lies not in eliminating human expertise but in expanding the range of tools available for interacting with increasingly large collections.

The enrichment activities undertaken within BelgicaWeb therefore demonstrate both the potential and the limitations of contemporary AI technologies. They offer promising directions for future development while simultaneously highlighting the need for careful governance and ongoing evaluation.

4.5.8 From Metadata to Research Infrastructure

The enrichment activities undertaken during BelgicaWeb suggest a broader transformation in how metadata can support access, discovery and reuse.

Historically, metadata often functioned primarily as a mechanism for organizing and retrieving resources. Within BelgicaWeb, metadata increasingly became viewed as a research resource in its own right. Quality indicators, geolocation data, named entities, summaries and other enrichment outputs provide additional layers of information that can support interpretation, exploration and analysis.

This shift has important implications for the future of web archiving. As collections continue to grow, the capacity to generate meaningful contextual information may become just as important as the ability to preserve content itself.

The enrichment infrastructure developed within BelgicaWeb provides an important step in this direction. By combining technical metadata, analytical metadata and AI-generated descriptions, the project sought to create richer pathways between preserved content and future users.

Key Findings and Recommendations

- Data enrichment plays a crucial role in transforming preserved content into reusable research resources.
- Researchers increasingly require contextual and analytical metadata in addition to traditional archival descriptions.
- Quality indicators can improve transparency and support critical interpretation of archived materials.
- Geographical enrichment introduces valuable additional perspectives for collection exploration and analysis.
- Copyright-related metadata can support more informed access management and rights assessment processes.
- Artificial intelligence offers significant opportunities for metadata generation but should be used in a transparent and responsible manner.
- Automated enrichment should complement rather than replace existing archival expertise.
- Metadata should increasingly be viewed as a research asset in its own right.
- Future developments should continue integrating enrichment, discovery and access within a single research-oriented infrastructure.

4.6 Building a Sustainable Legal and Ethical Framework for Access to Born-Digital Heritage

Throughout BelgicaWeb, legal and ethical considerations were not treated as external constraints imposed upon technical development. Instead, they gradually emerged as fundamental components of the project's overall access strategy. As collection creation, enrichment and platform development

progressed, it became increasingly clear that many practical decisions depended not only on technological feasibility but also on the legal and ethical frameworks within which cultural heritage institutions operate.

Questions that initially appeared primarily technical often revealed deeper legal dimensions. Decisions concerning what could be collected, how collections could be enriched, which datasets could be shared, whether researchers could perform text and data mining operations, or under what conditions APIs could expose archived content all required careful consideration of copyright, legal deposit, privacy law and emerging AI regulation. Consequently, the legal work undertaken during BelgicaWeb evolved from a supporting activity into a central pillar of the project's vision for sustainable access.

Rather than focusing exclusively on identifying legal barriers, the project sought to explore practical pathways through which cultural heritage institutions can responsibly preserve and provide access to born-digital collections. This approach reflects a broader understanding of sustainability. Long-term access depends not only on servers, storage and software, but also on governance structures capable of balancing openness, legal compliance, ethical responsibility and institutional mission.

4.6.1 The Evolving Legal Basis for Web Archiving

One of the major themes explored throughout the project concerned the legal basis for web archiving in Belgium. This question is of fundamental importance because collection building represents the foundation upon which all subsequent access activities depend. Without a sufficiently clear legal framework for acquisition and preservation, it becomes difficult to establish stable access services capable of supporting future research.

Historically, Belgian legal deposit legislation was developed within a publishing landscape dominated by printed works and physical carriers. While digital communication has become increasingly central to social, cultural and political life, the legislative framework has not always evolved at the same pace. This has created uncertainty regarding the legal position of born-digital heritage, particularly websites and social media content.

The legal analyses undertaken during BelgicaWeb coincided with a period of significant discussion concerning possible reforms to the legal deposit framework. The proposed changes envisage a substantial broadening of the categories of publications subject to legal deposit, including online publications and a range of born-digital materials. Such developments have potentially far-reaching implications for institutions seeking to preserve the Belgian web.

From the perspective of BelgicaWeb, the significance of these discussions extends beyond collection building alone. During the project, it became increasingly apparent that collection development and access are inseparable. Decisions regarding what may legally be collected directly influence what may ultimately be preserved, enriched and consulted. A stronger legal basis for acquisition therefore creates opportunities for more sustainable forms of access in the future.

At the same time, the project also demonstrated that legal deposit is not a complete solution to all challenges surrounding web archives. Even when materials can be collected and preserved, additional questions emerge concerning copyright, data protection, reuse and computational access. Sustainable access therefore depends upon a broader governance framework within which legal deposit forms one important component.

More generally, the discussions surrounding legal deposit highlighted the extent to which web archiving has become a matter of cultural policy. Decisions concerning what aspects of contemporary online culture are preserved today will influence what future generations are able to study tomorrow. BelgicaWeb therefore contributed not only to technical infrastructure development but also to broader discussions regarding the role of cultural heritage institutions in preserving digital society.

4.6.2 Copyright and the Challenge of Providing Access

If legal deposit provides the foundation for collection development, copyright remains one of the most significant factors shaping future access possibilities.

This challenge became increasingly visible as the project moved beyond harvesting and preservation towards questions of discovery, enrichment and reuse. Archived websites often contain a complex mixture of protected materials. A single webpage may combine textual content, photographs, audiovisual resources, software components, databases and user-generated contributions. Each of these elements may be subject to different rights frameworks and may involve different creators or rights holders.

As a result, copyright considerations affect virtually every stage of the archival lifecycle. They influence collection building, determine the conditions under which resources may be consulted and shape possibilities for computational analysis and dataset creation.

One of the most important findings emerging from the legal investigations is the need to distinguish clearly between preservation and access. The legal mechanisms that facilitate long-term preservation do not automatically create equivalent possibilities for unrestricted access or large-scale dissemination. Sustainable access therefore requires heritage institutions to navigate multiple legal frameworks simultaneously.

The project consequently explored a variety of copyright mechanisms potentially relevant to web archives. These included preservation exceptions for cultural heritage institutions, dedicated-terminal access, educational and research exceptions, out-of-commerce works provisions and text and data mining exceptions. Rather than identifying a single legal solution, the analyses demonstrated that access is likely to depend upon a combination of complementary approaches.

An important outcome of this work was the development of more operational perspectives on copyright compliance. Instead of treating copyright as a purely legal question, the project increasingly considered how copyright requirements might be translated into practical workflows, access policies and researcher agreements. This shift from abstract interpretation towards implementation represents one of the more significant contributions of the legal work package.

The investigations also reinforced the importance of flexibility. Different categories of material may require different access conditions, and future governance frameworks will likely need to accommodate a range of scenarios rather than applying a single uniform access model.

4.6.3 Research Access and Text and Data Mining

Among the legal topics examined during the project, text and data mining occupies a particularly important position because of its direct relevance to many of the research scenarios identified during consultations.

A recurring theme throughout BelgicaWeb was the growing expectation that web archives should support computational research. Researchers increasingly seek to analyze collections at scale using techniques such as text mining, network analysis, corpus analysis and machine learning. These methods depend upon access not only to individual webpages but also to larger datasets and structured corpora.

The project therefore devoted considerable attention to understanding the legal framework governing such activities. The analyses revealed that text and data mining exceptions create important opportunities for research while simultaneously introducing new responsibilities for both institutions and users.

Perhaps the most significant observation concerns the relationship between access and research. Computational research does not begin when analysis starts; it begins much earlier, with questions regarding who may access collections, under what conditions data may be shared and how research corpora can be created. Consequently, the legal framework for text and data mining cannot be viewed in isolation from broader access policies.

The project explored numerous practical scenarios involving requests for datasets, extracted links, social media collections and larger corpora. These examples highlighted the complexity of balancing researcher needs, copyright obligations and institutional responsibilities. In many situations, access decisions depend not on a single legal provision but on an assessment of multiple interacting factors.

Another recurring issue concerns the implementation of safeguards. Copyright exceptions often assume that institutions will maintain appropriate levels of security and governance. While legislation generally does not prescribe specific technical solutions, heritage institutions are nevertheless expected to develop mechanisms capable of preventing misuse and ensuring responsible access.

The project also highlighted continuing uncertainty surrounding opt-out mechanisms applicable to certain forms of text and data mining. The absence of harmonized standards creates practical challenges for institutions managing large and complex collections. This area remains one in which technological developments continue to outpace legal and regulatory clarity.

Taken together, the analyses suggest that text and data mining is likely to become one of the most important future use cases for web archives. Supporting such activities will require not only technical infrastructure but also carefully designed governance frameworks capable of balancing openness and responsibility.

4.6.4 Personal Data Protection and Responsible Access

Alongside copyright, personal data protection emerged as one of the most significant legal considerations affecting future access to born-digital heritage.

Web archives inevitably preserve traces of human activity. Websites, blogs, forums and social media platforms frequently contain information relating to identifiable individuals. Consequently, efforts to preserve contemporary digital culture must also account for the rights and interests of the people whose data may appear within archived collections.

The project built upon previous Belgian research in this area while examining more recent developments within European data protection discussions. Particular attention was devoted to the relationship between archival preservation, scientific research and personal data processing.

An important observation emerging from this work concerns the growing recognition of scientific research as a socially valuable activity. Recent discussions at the European level increasingly seek to clarify the legal framework surrounding research access and the conditions under which scientific projects may benefit from specific provisions or safeguards. These developments are highly relevant for BelgicaWeb because many future users of web archives are expected to be researchers operating within academic environments.

At the same time, the project repeatedly emphasised that scientific value alone cannot justify unrestricted access. Sustainable access requires a careful balance between research potential and the protection of individual rights. Transparency, proportionality and appropriate safeguards therefore remain central considerations.

This balance is particularly important because many born-digital collections contain information that was originally created for purposes other than archival preservation. The project therefore highlighted the need for governance structures capable of evaluating requests in context and considering both research objectives and potential risks.

Viewed more broadly, the discussions surrounding personal data protection reinforce one of the central themes running throughout BelgicaWeb: access should not be measured solely by how much information can be made available, but also by how responsibly that information is managed.

4.6.5 Artificial Intelligence, Governance and Institutional Responsibility

Legal and ethical discussions surrounding artificial intelligence became increasingly relevant as the project explored the use of AI-supported enrichment techniques.

The analyses undertaken during BelgicaWeb suggest that many of the AI applications under consideration, including summarization, named entity recognition, sentiment analysis and content categorization, are unlikely to fall within the highest-risk categories of current regulatory frameworks. Nevertheless, the use of such technologies introduces important questions regarding accountability, transparency and institutional preparedness.

One of the most significant observations concerns the relationship between automation and expertise. Throughout the project, AI was consistently viewed as a tool for supporting human

interpretation rather than replacing it. Automated enrichment may help researchers navigate complex collections, but its outputs cannot be treated as definitive representations of reality.

This perspective influenced the project's broader approach to AI governance. Rather than viewing compliance as a purely legal exercise, BelgicaWeb increasingly focused on organizational capacity. Institutions need staff capable of understanding how AI systems function, recognizing their limitations and evaluating the reliability of their outputs.

The growing importance of AI literacy therefore emerged as a recurring theme. Future access infrastructures will increasingly depend upon a combination of technological innovation and human expertise. Ensuring that cultural heritage professionals possess the skills required to interact with these systems responsibly may prove just as important as the technologies themselves.

More generally, the AI-related activities undertaken during BelgicaWeb illustrate how rapidly changing technological environments can create new governance challenges. Sustainable access requires institutions not only to adopt new technologies but also to develop frameworks capable of ensuring that these technologies remain transparent, trustworthy and aligned with public values.

4.6.6 Openness, Licensing and FAIR Reuse

A final theme explored throughout the project concerns openness and reuse.

One of the guiding principles of BelgicaWeb was the aspiration to make born-digital heritage as findable, accessible, interoperable and reusable as possible. However, the project quickly demonstrated that openness is not simply a technical matter. It is equally shaped by legal, organizational and strategic decisions.

Questions concerning licensing therefore became increasingly important. The project examined different forms of open-source software licensing, open-content licensing and rights statements in order to better understand how institutions can communicate access and reuse conditions clearly.

An important lesson emerging from this work is that openness exists on a spectrum rather than as a simple binary distinction. Collections may be publicly discoverable while access to content remains restricted. Metadata may be openly reusable while underlying content remains protected. Different categories of material may require different approaches depending on legal, ethical and practical considerations.

The project also highlighted the importance of rights communication. Researchers benefit significantly when the conditions governing access and reuse are expressed clearly and consistently. Standardised rights statements and licensing frameworks therefore have the potential to reduce uncertainty and support more effective reuse of preserved collections and datasets.

At a broader level, the discussions surrounding licensing reinforce the interconnected nature of the project's various activities. Collection development, access, enrichment and legal compliance all ultimately influence the extent to which born-digital heritage can contribute to future research and cultural understanding.

4.6.7 Towards Responsible and Sustainable Access

Taken together, the legal and ethical investigations undertaken during BelgicaWeb demonstrate that sustainable access depends upon much more than technological capability.

Web archives operate within a complex environment shaped by copyright law, data protection requirements, institutional mandates, public expectations and rapidly evolving technological developments. The challenge is therefore not simply to maximize access but to create forms of access that remain legally robust, ethically responsible and institutionally sustainable over time.

One of the central achievements of BelgicaWeb lies in demonstrating that these objectives do not necessarily conflict with one another. Through careful governance, transparent policies, researcher engagement and thoughtful technical design, it is possible to create infrastructures that simultaneously support preservation, research and responsible access.

Rather than presenting legal requirements as obstacles, the project increasingly viewed them as design parameters that help shape sustainable access strategies. In doing so, BelgicaWeb contributes not only to the preservation of born-digital heritage but also to the development of governance models capable of supporting its long-term use.

Key Findings and Recommendations

- Sustainable access depends on the interaction of legal, technical, organisational and ethical considerations.
- Legal deposit reform has the potential to strengthen the long-term foundations of Belgian web archiving.
- Copyright remains one of the most important factors shaping future access and reuse possibilities.
- Text and data mining is likely to become a central research use case and should be incorporated into future access strategies.
- Personal data protection requires careful balancing of research value and individual rights.
- AI governance should focus not only on compliance but also on transparency, accountability and institutional capacity building.
- Licensing and rights communication play an important role in enabling FAIR-oriented reuse.
- Researcher agreements and governance frameworks can help translate legal requirements into practical access workflows.
- Legal and ethical considerations should be integrated throughout the archival lifecycle rather than addressed only at the point of access.
- Responsible governance is an essential component of sustainable access to born-digital heritage.

4.7 Sustainability, Governance and Future Perspectives

Sustainability emerged as a cross-cutting theme throughout BelgicaWeb. The project demonstrated that long-term access to born-digital heritage depends not only on preservation technologies but also

on governance structures, institutional commitment, legal certainty and ongoing engagement with user communities.

A key principle throughout the project was the use of open and interoperable technologies. Components such as Browsertrix, SolrWayback and Symfony reduce dependence on proprietary solutions and facilitate collaboration with the international web archiving community. However, technical openness alone is insufficient. Sustainable infrastructures also require documentation, maintenance capacity and long-term organizational support.

The project further demonstrated the importance of interdisciplinary collaboration. Sustainable access to born-digital heritage requires expertise spanning digital preservation, software development, legal analysis, metadata management and research support. Future development should therefore remain embedded within institutional structures rather than relying solely on project-based activities.

The Expert Panel proved particularly valuable as a mechanism for continuous dialogue with researchers and heritage professionals. Beyond requirements gathering, its contributions improved collection development, quality assurance and service design. Similar participatory approaches should continue to be integrated into future initiatives.

BelgicaWeb also highlighted the growing role of web archives as research infrastructures. User consultations revealed increasing demand for computational access, reusable datasets and advanced analytical services. Future developments should therefore continue to strengthen interoperability, enrichment services and integration with wider research ecosystems.

Finally, the project underscored the importance of adaptability. Legal frameworks, online platforms and digital technologies continue to evolve rapidly. Sustainable infrastructures must therefore remain sufficiently flexible to accommodate future developments while maintaining responsible and legally compliant access.

Rather than producing a single standalone system, BelgicaWeb established the foundations required for sustainable access to born-digital heritage in Belgium. Its main contribution lies in bringing together user research, collection development, access infrastructure, data enrichment and legal governance within a single integrated framework. This holistic approach provides a roadmap for future development and demonstrates how born-digital heritage can evolve from a preservation challenge into a sustainable research infrastructure.

Key Findings and Recommendations

- Sustainability depends on technical, organisational, legal and social factors.
- Open and interoperable technologies provide an important foundation for long-term development.
- Sustainable access requires ongoing institutional commitment beyond project funding.
- Interdisciplinary collaboration remains essential.
- User communities should continue to participate in collection development and service evaluation.
- Web archives should be developed as research infrastructures that support discovery, analysis and reuse.

- Future development should prioritise flexibility and adaptability.
- BelgicaWeb provides a strong foundation for future Belgian web and social media archiving initiatives.

4.8 Synthesis: Towards Sustainable Access to Belgium's Born-Digital Heritage

One of the central ambitions of BelgicaWeb was to move beyond the question of how born-digital heritage can be collected and preserved and instead address the broader challenge of how such heritage can remain accessible, usable and meaningful over time. Throughout the project it became increasingly clear that sustainable access cannot be reduced to a single technical solution or organizational intervention. Rather, it emerges from the interaction of multiple dimensions, each of which contributes to the long-term viability of born-digital heritage infrastructures.

The project demonstrated that sustainable access begins with an understanding of users and their evolving research practices. The workshops, Expert Panel consultations and international survey all pointed towards the same conclusion: researchers increasingly expect web archives to function as research infrastructures rather than passive repositories. Future users seek not only the ability to consult archived webpages but also mechanisms that support exploration, dataset creation, computational analysis and long-term reuse. These findings shaped many of the subsequent decisions taken throughout the project and reinforced the importance of a user-centred approach to infrastructure development.

The collection development activities highlighted the importance of combining technical harvesting capabilities with transparent and inclusive curation processes. Through the creation of seed lists, the involvement of the Expert Panel and the evaluation of harvesting infrastructures, the project demonstrated that collection building is not simply a question of capture technology. Decisions regarding what to preserve and how to document these decisions have a direct impact on future research possibilities. Sustainable access therefore begins at the point of selection and depends upon creating collections that are both technically robust and intellectually meaningful.

The development of the access platform and accompanying API explored how archived content can be transformed into a genuinely usable research environment. The project recognized that future users will interact with collections in different ways and therefore require multiple forms of access. While graphical interfaces remain important for discovery and exploration, APIs and export mechanisms are increasingly essential for computational research. By combining these approaches, BelgicaWeb laid the foundations for an access model capable of accommodating diverse user communities and research methodologies.

The enrichment activities demonstrated the growing importance of metadata as a bridge between preserved content and future use. Quality indicators, geolocation information, copyright markers and AI-generated descriptions all contribute additional layers of knowledge that help users understand and navigate large collections. The project showed that metadata should be viewed not merely as an aid to retrieval but increasingly as a research resource in its own right. This perspective becomes particularly important as collections continue to grow and manual exploration becomes less feasible.

The legal and ethical analyses undertaken throughout the project confirmed that sustainable access depends upon governance as much as technology. Questions concerning legal deposit, copyright, personal data protection, text and data mining and artificial intelligence repeatedly influenced decisions regarding collection development and access. Rather than treating these issues as obstacles, BelgicaWeb demonstrated that they can serve as design parameters guiding the development of responsible and sustainable infrastructures. The translation of legal principles into practical workflows, researcher agreements and governance frameworks represents an important contribution of the project.

Perhaps one of the most significant findings emerging from BelgicaWeb concerns the value of interdisciplinary collaboration. The project brought together expertise from web archiving, digital preservation, computer science, law and research support. Throughout the project, it became increasingly apparent that none of these perspectives alone would be sufficient to address the challenges associated with born-digital heritage. Sustainable access requires the continuous interaction of technical expertise, legal knowledge, curatorial judgement and user engagement.

The role of the Expert Panel provides a particularly strong illustration of this principle. Through their contributions to seed list creation, user requirements analysis, enrichment discussions and quality assurance activities, panel members helped ensure that project developments remained connected to the needs and expectations of future users. Their involvement demonstrated that access infrastructures benefit substantially from ongoing dialogue with the communities they intend to serve.

Taken together, the outcomes of BelgicaWeb suggest that the future of web archiving lies not merely in preservation but in the creation of integrated research environments that support discovery, analysis and reuse. Such environments require collections, access mechanisms, enrichment workflows and governance structures that operate in concert rather than as isolated components. The project therefore contributes to a broader shift in how born-digital heritage is conceptualized: from a preservation challenge towards a sustainable research infrastructure capable of supporting future scholarship, cultural memory and public understanding.

Ultimately, BelgicaWeb does not provide a final answer to the challenges associated with born-digital heritage. The technological, legal and social environments within which web archives operate will continue to evolve, and future infrastructures will need to adapt accordingly. What the project does provide is a foundation upon which such future developments can build. By integrating user-centered design, collection development, access innovation, enrichment technologies and responsible governance, BelgicaWeb establishes an important framework for sustaining access to Belgium's born-digital heritage and contributes valuable insights to the wider international web archiving community.

Key Findings and Recommendations

- Sustainable access should be understood as a multidisciplinary challenge requiring the integration of technical, legal, organizational and user-oriented perspectives.
- Collection development, access and governance are closely interconnected and should be addressed holistically.
- Web archives are increasingly evolving into research infrastructures that support discovery, analysis and reuse.
- Metadata enrichment significantly increases the research value of born-digital collections.
- APIs and computational access mechanisms are becoming essential components of contemporary archival infrastructures.
- Legal and ethical considerations should be integrated throughout the archival lifecycle rather than addressed after collections have been created.
- Continuous engagement with researchers and other stakeholders improves both relevance and usability.
- Open and interoperable technologies provide important foundations for long-term sustainability.
- Future development efforts should continue to balance innovation, transparency and responsible governance.
- The methodologies, infrastructure components and lessons learned within BelgicaWeb provide a strong basis for future web and social media archiving initiatives in Belgium and beyond.

Principal Project Outputs

The BelgicaWeb project produced a set of complementary outputs that together contribute to sustainable access to Belgium's born-digital heritage. These outputs include:

- A user requirements study based on workshops, surveys and Expert Panel consultations.
- Seed lists, collection development methodologies and harvesting workflows for born-digital heritage.
- Experimental web archive collections created using Browsertrix.
- A prototype access platform designed for the discovery and consultation of archived web content.
- An API supporting programmatic access and future computational research applications.
- Metadata enrichment workflows incorporating quality assessment, geolocation, copyright indicators and AI-assisted enrichment techniques.
- A legal and ethical framework addressing copyright, data protection, text and data mining, AI regulation and reuse.
- Sustainability and governance recommendations supporting the long-term development of Belgian born-digital heritage infrastructures.

Together, these outputs provide both practical tools and strategic guidance for the future preservation, access and reuse of web and social media archives in Belgium.

5. DISSEMINATION AND VALORISATION

5.1 Conferences attended

- IIPC Web Archiving Conference, Paris, France, 23-26 April 2024.
- DH Benelux Conference, Leuven, Belgium, 4-6 June 2024.
- iPRES 2024 Conference, Ghent, Belgium, 19 September 2024.
- Event Praktijknetwerk Sociale Media Archivering, The Hague, The Netherlands, 15 October 2024.
- Born-Digital Collections, Archives and Memory Conference, London, United Kingdom, 2-4 April 2025.
- IIPC Web Archiving Conference, Oslo, Norway, 8-10 April 2025.
- RESAW Conference, Siegen, Germany, 4-6 June 2025.
- Cycle de rendez-vous francophones en ligne d’IIPC, online, September-November 2025.

5.2 Presentations and workshops organized

Throughout the project, consortium members organised and delivered presentations, workshops, demonstrations and panel contributions at national and international events. These activities served multiple purposes: gathering user requirements, validating project results, strengthening stakeholder networks and disseminating emerging findings to research and heritage communities.

Particular attention was devoted to engagement with communities active in web archiving, digital humanities, social media archiving and digital preservation. Workshops organised at DH Benelux and within the Dutch-Flemish Network of Practice for Social Media Archiving directly contributed to the project's user-centred methodology by informing requirements gathering and co-creation activities.

The project team also presented BelgicaWeb results at major international forums, including IIPC, RESAW, iPRES and the Born-Digital Collections, Archives and Memory conference. These activities increased the visibility of Belgian born-digital heritage initiatives and facilitated international exchange on access, preservation, legal governance and metadata enrichment.

A complete list of presentations and conference contributions is included in Section 6.2.

5.3 Project Websites

BelgicaWeb maintained project webpages through KBR and Ghent University. These platforms were used to communicate project objectives, share updates on ongoing activities, promote surveys, workshops and events, and provide access to information about project outputs and dissemination activities.

5.4 Participation in professional networks

International Internet Preservation Consortium (IIPC)

KBR participates actively in the International Internet Preservation Consortium (IIPC). BelgicaWeb developments were regularly communicated through annual conferences, working groups and regional webinar activities.

Particular attention was devoted to the French-speaking regional activities of IIPC, where project members presented both legal and technical results of BelgicaWeb.

Praktijknetwerk voor Sociale Media Archivering

KBR remained actively involved in the Dutch-Flemish Praktijknetwerk voor Sociale Media Archivering. Through this network, project members exchanged expertise concerning replay technologies, social media archiving workflows and access solutions.

The network also served as an important channel for obtaining feedback on BelgicaWeb developments from peer institutions.

5.5 Internal dissemination and institutional impact

BelgicaWeb generated benefits that extend beyond the immediate project outputs. At KBR, the project contributed to the development of expertise in web and social media archiving, harvesting workflows, access infrastructures and metadata enrichment. It also strengthened collaboration between collection specialists, ICT staff and researchers, helping to embed born-digital heritage more firmly within future planning and service development.

At Ghent University and the University of Namur, the project further strengthened interdisciplinary collaboration between digital humanities, computer science and legal research communities. The expertise, methodologies and professional networks developed during BelgicaWeb provide a basis for future research projects, infrastructure development and international collaboration in the field of born-digital heritage.

KBR

Project updates were regularly presented during meetings of the Operational Directorate of Contemporary Collections. Demonstrations and exchanges were also organised with KBR's ICT department to discuss infrastructure requirements and future sustainability.

UGent

At IDLab, project requirements informed the training trajectory of the doctoral researcher working on BelgicaWeb, particularly in the areas of knowledge graphs and large language models.

At MICT and GhentCDH, project developments were regularly shared with colleagues to ensure alignment and encourage interdisciplinary collaboration.

UNamur

Within CRIDS, project developments and legal analyses were regularly discussed among colleagues working on digital regulation, privacy and intellectual property issues.

5.6 Open science and repositories

SODHA

The project deposited research data concerning the user requirements study and expert panel activities in the Social Sciences and Digital Humanities Archive ([SODHA](#)).

The most important dataset produced during the project consists of the repository containing survey materials, workshop materials, metadata and results relating to the user requirements study.

University repositories

Research outputs were also disseminated through the institutional repositories of the participating universities, including [PURE](#) (UNamur) and [Biblio](#) (UGent).

Orfeo

Reports produced within the project were also deposited on [Orfeo](#), the Open Repository for Federal Organisations. A full list can be found in point 6.4 Technical and Research reports underneath.

Github

In January 2027 we plan to publish the BelgicaWeb code on GitHub as open source on this URL: <https://github.com/kbrbe/belgicaweb>.

5.7 Public communication and outreach

Invited lectures

Geeraert, F. & Vandendyck, C. (2026). [WinIT 25-26], Executive Master in Digital & IT Essentials, ULB, 6 March 2026.

Geeraert F., (2026). [Webarchivering], Master in Digitaal Erfgoed, geschiedenis en publiek, UGent, 24 March 2026.

Social media

The user requirements survey and project updates were disseminated through the communication channels of KBR and UGent.

Newsletters

Project activities and surveys were promoted through institutional newsletters, including the KBR newsletter.

Press coverage

The project contributed to broader public discussions concerning the preservation of online heritage and born-digital collections.

Notable media coverage included:

- *La course contre la montre pour protéger le web contre Donald Trump*, Le Soir, 9 February 2025.

5.8 Concluding symposium

The BelgicaWeb concluding symposium took place on 20 April 2026 at KBR in Brussels.

During the symposium, the project team presented the main outcomes of BelgicaWeb, including the user requirements study, collection development activities, the access platform and API, metadata enrichment workflows, legal analyses and future sustainability plans. The programme opened with the plenary presentation *Digital Heritage for the Future: the BelgicaWeb Story*, chaired by Julie Birkholz and presented by Julie Birkholz, Christina Vandendyck, Peter Mechant, George Caique Gouveia Barbosa, Dieter De Witte, Élodie Lecroart and Friedel Geeraert.

The symposium also featured three thematic panel discussions addressing user-centred design of web archives, the legal challenges associated with web archiving and open collections, and the role of artificial intelligence in web and social media archives. Together, these sessions provided an international perspective on the future development of web archiving infrastructures and access services

The symposium brought together researchers, cultural heritage professionals, web archivists, legal scholars, digital humanities specialists and policy stakeholders. Organised in conjunction with the IIPC 2026 conference and the Early Scholars Spring School on Web Archives, it served as the principal dissemination event marking the conclusion of the BelgicaWeb project and provided an opportunity to situate its results within ongoing international discussions on web and social media archiving.

The symposium featured the following panel discussions:

- *Users First: (Re)Designing Web Archives Around Real Needs*, chaired by Peter Mechant, with Sara Aubry, Eveline Vlassenroot, Jon Carlstedt Tønnessen, Andrea Kocsis and Natasha Kitcher.
- *The Legal Challenges of Web Archiving and Open Collections*, chaired by Élodie Lecroart, with Séverine Dusollier, Stef van Gompel and Frédéric Young.
- *AI in Web and Social Media Archives*, chaired by Julie Birkholz, with Jefferson Bailey, Dorothée Benhamou-Suesser, Victor Harbo Johnston and Hannes Lowagie.

6. PUBLICATIONS

6.1 Peer-reviewed journal articles

- Vlassenroot, E., Tao, Y. & Mechant, P. (2025). *The evolving landscape of web and social media archiving: a comprehensive review of current practices in national libraries and archives*. International Journal of Digital Humanities. <https://doi.org/10.1007/s42803-025-00106-8>

6.2 Conference presentations and proceedings

DH Benelux

- Vlassenroot, E., Tao, Y., Mechant, P., Vandendyck, C. & Geeraert, F. (2024). *Opening up born-digital resources*. Workshop presentation, DH Benelux Conference. <https://2024.dhbenelux.eu/>

IIPC

- Geeraert, F. (2024). *Web Archiving Update from KBR – Belgium*. Presentation, IIPC Web Archiving Conference. <https://netpreserve.org/ga2024/>
- Vandendyck, C. (2025). *Enhancing Accessibility to Belgian Born-Digital Heritage: The BelgicaWeb Project*. Presentation, IIPC Web Archiving Conference. <https://netpreserve.org/ga2025/>
- Vlassenroot, E., Mechant, P., Geeraert, F. & Vandendyck, C. (2025). *Beyond Preservation: Engaging Audiences and Researchers with Web Archives*. Presentation, IIPC Web Archiving Conference. <https://netpreserve.org/ga2025/>

Born-Digital Collections, Archives and Memory

- Vandendyck, C. (2025). *FAIR Principles in Action: Archiving and Accessing Belgium’s Born-Digital Heritage*. <https://www.sas.ac.uk/news-events/events/born-digital-collections-archives-memory-0>

RESAW

- Vandendyck, C. (2025). *Demonstration of BelgicaWeb: Sustaining Access to Belgium’s Born-Digital Heritage*. <https://www.mediacoop.uni-siegen.de/datafiedweb/>
- Vlassenroot, E. (2025). *Navigating the Datafied Web: User Requirements and Literacy with Web Archives*. <https://www.mediacoop.uni-siegen.de/datafiedweb/>

IIPC French-speaking webinar series

- Lecroart, E. (2025). *Accès, utilisation et extraction des (méta)données issues des archives du web: enjeux techniques et juridiques*. <https://netpreserve.org/event/iipc-regional-series-research-services-and-tools-for-accessing-web-archives-giving-collections-access/>

- Vandendyck, C. (2025). *BelgicaWeb: Préserver le patrimoine numérique belge. Recherche, services et outils d'accès aux archives web*. <https://netpreserve.org/event/iipc-regional-series-research-services-and-tools-for-accessing-web-archives-giving-collections-access/>

iPRES

- Kávvdia, Z., Geeraert, F., Van Keer, E. & Weyns, K. (2024). *Surmounting the barriers to social media archiving with a little help from my friends*. Birds of a feather session, iPRES 2024, Ghent. <https://ipres2024.pubpub.org/>

6.3 Datasets

SODHA datasets

- Vandendyck, C. & Geeraert, F. (2026). *Report BelgicaWeb: Data Collection and Description of Process* (Version 1). Social Sciences and Digital Humanities Archive (SODHA). DOI: <https://doi.org/10.34934/DVN/0VP9I1>.
- Geeraert, F. & Vandendyck, C. (2025). *Seed Lists for the BelgicaWeb Research Project Related to the Archiving of Web and Social Media Content*. Social Sciences and Digital Humanities Archive (SODHA). DOI: <https://doi.org/10.34934/DVN/CODGSS>.
- Mechant, P. (2025). *Survey on User Requirements for Web Archiving*. Social Sciences and Digital Humanities Archive (SODHA). DOI: <https://doi.org/10.34934/DVN/NHLLHP>.

6.4 Technical and research reports

The BelgicaWeb project produced a range of reports and deliverables, including:

- Selection Policy for Belgian Born-Digital Heritage Collections (D1.2.1) - internal document
- BelgicaWeb Seed List (D1.2.2, versions 1 and 2).
 - Geeraert, Friedel; Vandendyck, Christina, 2025, "Seed lists for BelgicaWeb research project related to the archiving of web and social media content", <https://doi.org/10.34934/DVN/CODGSS>, Social Sciences and Digital Humanities Archive – SODHA.
- List of Expert Panel Members and Meetings (D1.1.1) - internal document
- Status Report on the Expert Panel and Stakeholder Engagement Activities (D1.1.2) - internal document
- Collections and Data Capture Process Documentation (D1.3.1)
 - Vandendyck, Christina; Geeraert, Friedel, 2026, "Report belgicaweb - Data collection & description of process", <https://doi.org/10.34934/DVN/0VP9I1>, Social Sciences and Digital Humanities Archive – SODHA.
- Quality Assurance Report (D1.4.1) - internal document

- Technical and Functional Requirements for Web and Social Media Archives (D2.1.1 & D3.1.1).
 - Mechant, Peter, 2025, "Survey on user requirements for web archiving", <https://doi.org/10.34934/DVN/NHLLHP>, Social Sciences and Digital Humanities Archive – SODHA.
- Indexes and Documentation of the BelgicaWeb Indexing Process (D2.2.1)
 - Vandendyck, Christina, 2026, “BelgicaWeb research project Deliverable 2.2.1 Indexes and documentation about the indexing process”, <https://orfeo.belnet.be/handle/internal/15254>, BELSPO Institutional Open Access Repository for Federal Organisations – orfeo.
- API, Access Platform and User Documentation (D2.3.1) - version 1 published on SODHA
 - Vandendyck, Christina, 2026, "BelgicaWeb research project Deliverable 2.3.1 API, access platform, and user guides and documentation v01", <https://doi.org/10.34934/DVN/SV2PPK>, Social Sciences and Digital Humanities Archive – SODHA.
- Legal Framework for Data Exchange, Text and Data Mining and Opening Up Collections (D2.4.1).
 - Lecroart, Elodie & Michel, Alejandra, 2026, “BelgicaWeb Research Project - Deliverable 2.4.1 Analysis of the legal framework allowing data exchange, text and data mining and the opening up of collections”, <https://orfeo.belnet.be/handle/internal/15258>, BELSPO Institutional Open Access Repository for Federal Organisations – orfeo.
- Enriched Datasets and Enrichment Workflow Documentation (D3.2.1) - internal document
- Data Protection, Privacy Rights, Freedom of Expression and the European Artificial Intelligence Act (D3.3.1).
 - Lecroart, Elodie & Michel, Alejandra, 2026, “BelgicaWeb Research Project Deliverable 3.3.1 Report on Personal Data Protection”, <https://orfeo.belnet.be/handle/internal/15261>, BELSPO Institutional Open Access Repository for Federal Organisations – orfeo.
 - Lecroart, Elodie & Michel, Alejandra, 2026, “BelgicaWeb Research Project - Deliverable 3.3.1 Report on Artificial Intelligence Act (AI Act)”, <https://orfeo.belnet.be/handle/internal/15259>, BELSPO Institutional Open Access Repository for Federal Organisations – orfeo.
 - Lecroart, Elodie & Michel, Alejandra, 2026, “BelgicaWeb Research Project - Synthetic report on free/open-source licences”, <https://orfeo.belnet.be/handle/internal/15260>, BELSPO Institutional Open Access Repository for Federal Organisations – orfeo.

- Frequently Asked Questions on Legal Issues Related to Web and Social Media Archiving (D4.2.1) - internal document
- Sustainability and Institutional Embedding Report (D5.1.1)
 - Geeraert, Friedel; Vandendyck, Christina, 2026, "BelgicaWeb research project Deliverable 5.1.1 Policy document reflecting on governance and sustainability and operational procedures", <https://doi.org/10.34934/DVN/YODIFX>, Social Sciences and Digital Humanities Archive – SODHA.
- Action Plan for Valorisation (D7.1.2, versions 1-3) - internal document
- Data Management Plan (D8.1.1, revised editions) - internal document

6.5 Professional publications and blog posts

- Geeraert, F. (2024). *Reflections on the 2024 IIPC General Assembly and Web Archiving Conference*. <https://netpreserveblog.wordpress.com/2024/05/08/reflections-on-the-2024-iipc-general-assembly-and-web-archiving-conference/>
- Geeraert, F. & Vandendyck, C. (2025). *Reflections on the 2025 IIPC General Assembly and Web Archiving Conference*. <https://netpreserveblog.wordpress.com/2025/08/29/reflections-on-the-2025-iipc-general-assembly-and-web-archiving-conference/>

7. ACKNOWLEDGEMENTS

The consortium would like to thank all members of the scientific follow-up committee, expert panel members, workshop participants, survey respondents and interviewees who contributed their expertise throughout the project.

Special thanks are due to the partner institutions KBR, Ghent University and the University of Namur for their collaboration and support.

The consortium would also like to thank colleagues, technical staff and external collaborators who contributed to infrastructure development, data collection, dissemination and project administration.

Finally, we gratefully acknowledge the Belgian Science Policy Office (BELSPO) for funding the BelgicaWeb project and supporting interdisciplinary research into Belgium's born-digital heritage and the valorisation thereof.